# Preservación de Consultas Expresables en Fragmentos Existenciales en Bases de Datos Relacionales

Alejandro L. Grosso
Dpto. Informática
Universidad Nacional de San Luis
San Luis, Argentina.
**e-mail**:agrosso@unsl.edu.ar

J.M. Turull-Torres
Information Science Research Centre
Information Systems Dept.
Massey University
Wellington, New Zealand.
**e-mail**:J.M.Turull@massey.ac.nz

**Resumen**

En el contexto de bases de datos, el concepto de preservación de teorías en una cierta lógica y entre dos estructuras relacionales $\mathcal{A}$ y $\mathcal{B}$ dadas, significa que el conjunto de consultas Booleanas que son expresables en dicha lógica y que son verdaderas en la base de datos $\mathcal{A}$ también son verdaderas en la base de datos $\mathcal{B}$. Así, por ejemplo ante el agregado de nuevos elementos a una base de datos es posible determinar si el conjunto de consultas Booleanas expresables en $FO^k(\exists)$ (y en la extensión infinitaria $L^k_{\infty,\omega}(\exists)$) se preserva. Existen juegos de fichas que caracterizan la preservación de teorías para estas dos lógicas, y es sabido que dicha preservación puede determinarse en tiempo polinomial. En el presente trabajo mostramos una caracterización alternativa de la preservación de teorías en $FO^k(\exists)$ (y en la extensión infinitaria $L^k_{\infty,\omega}(\exists)$) mediante la realización de tipos de $k$-tuplas. Luego presentamos un algoritmo polinomial que permite determinar dicha preservación y en el que utilizamos nuestra caracterización.

**Palabras Claves**: Teoría de Modelos Finitos, Preservación de Teorías, Bases de Datos Relacionales.

# 1. Introducción

Desde la perspectiva de teoría de modelos finitos una base de datos es un conjunto finito de objetos con distintas relaciones finitas entre ellos y una lógica es un lenguaje, entonces podemos usar esa lógica para describir cómo es que ese conjunto de objetos se encuentra estructurado. Las sentencias de una lógica expresan consultas Booleanas. Así dada una lógica $\mathcal{L}$, el conjunto de todas las consultas Booleanas expreables en $\mathcal{L}$ que se satisfacen en la base de datos conforma la $\mathcal{L}$-teoría de la base de datos. Decimos que la $\mathcal{L}$-teoría de la base de datos $\mathcal{A}$ se preserva en la base de datos $\mathcal{B}$ si el conjunto de consultas Booleanas que se satisfacen en $\mathcal{A}$ y que son expresables en $\mathcal{L}$ está incluido en en el conjunto de consultas Booleanas que se satisfacen en $\mathcal{B}$ y que son expresables en $\mathcal{L}$. Otro concepto importante en teoría de modelos finitos es la noción de *tipos*. Ésta consiste en considerar todas las propiedades que posee una $k$-tupla en la base de datos. El conjunto de fórmulas con a lo sumo $k$-variables libres que se satisfacen para dicha $k$-tupla conforma el tipo de la $k$-tupla. Intuitivamente el tipo de una $k$-tupla representa la clase de consultas para las cuales la $k$-tupla o algunas de sus subtuplas aparece en la respuesta de la misma.

Así podemos definir relaciones entre $k$-tuplas dependiendo de la relación que existe entre los tipos de las mismas. Si dos $k$-tuplas, $\bar{a}$ y $\bar{b}$ poseen el mismo $\mathcal{L}$-tipo significa que no las podemos distinguir en $\mathcal{L}$, es decir, o bien aparecen las dos en el resultado de una consulta en $\mathcal{L}$ o no aparece ninguna de las dos. En el caso en que el $\mathcal{L}$-tipo de la $k$-tupla $\bar{a}$ esté incluido en $\mathcal{L}$-tipo de la $k$-tupla $\bar{b}$, significa que, si la $k$-tupla $\bar{a}$ aparece en el resultado de una consulta también aparecerá la $k$-tupla $\bar{b}$. Y si la $k$-tupla $\bar{b}$ no aparece en el resultado de una consulta lo mismo sucederá con la $k$-tupla $\bar{a}$. De esta manera los $\mathcal{L}$-tipos representan los conjuntos de tuplas a partir de los cuales se construye la respuesta a una consulta expresada en la lógica $\mathcal{L}$.

En teoría de modelos finitos existe una caracterización de la preservación de teorías para el fragmento $FO^k(\exists)$ (y el fragmento infinitario $L^k_{\infty,\omega}(\exists)$) debida a Kolaitis y Vardi [5]. Tal caracterización es realizada en base a una generalización de los juegos de fichas debidos a Ehrenfeucht-Fraïssé.

Nosotros damos en la proposición 4.3 una caracterización alternativa [4] en base a los $FO^k(\exists)$-tipos realizados en las bases de datos. Y en la sección 4.1 proponemos un algoritmo polinomial que verifica la satisfacción de la prop. 4.3.

# 2. Preliminares

A menos que sea establecido, en el presente trabajo nosotros seguiremos la notación usual en teoría de modelos, como en [3].

## 2.1. Sintaxis de la Lógica de Primer Orden

Desde la perspectiva de la teoría de modelos las lógicas son consideradas lenguajes definidos a partir de un alfabeto de símbolos que incluye un vocabulario. Así una lógica la denotamos como $\mathcal{L}_\sigma$ donde $\sigma$ es un vocabulario. En nuestro caso nos limitaremos a vocabularios puramente relacionales. Un *vocabulario* es un conjunto finito y no vacío de símbolos de relación cada uno asociado con un número natural que representa su aridad. Por ejemplo la lógica de primer orden ($FO$) sobre un vocabulario relacional consiste de:

1) Un vocabulario relacional $\sigma =< R_1, \ldots, R_s >$ con aridades $r_1, \ldots, r_s$ respectivamente; 2) Un conjunto de variables $X = \{x_1, \ldots, x_n, \ldots\}$; 3) Un conjunto de cuantificadores $Q = \{\exists\}$; 4) Un conjunto de conectivas $C = \{\neg, \wedge\}$; 5) Un alfabeto $\{=, (, )\} \cup \sigma \cup X \cup Q \cup C$; 6) Un conjunto de términos $Term = \{x_1, \ldots, x_n, \ldots\}$; 7) Un conjunto de fórmulas atómicas $F_0 = \{t_i = t_j : t_i, t_j \in Term\} \cup \{R_i(t_1, \ldots, t_{r_i}) : t_1, \ldots, t_{r_i} \in Term$ y $R_i \in \sigma$ con aridad $r_i\}$; 8) $F_{n+1} = F_n \cup \{(\neg\varphi) : \varphi \in F_n\} \cup \{(\varphi \wedge \psi) : \varphi, \psi \in F_n\} \cup \{(\exists x_i \varphi) : \varphi \in F_n\}$; 9) El conjunto de fórmulas bien formadas $\mathcal{F} = \bigcup_{n \in \omega} F_n$.

Si $\varphi$ es una fórmula atómica todas las variables de $\varphi$ están en $free(\varphi)$; $free(\neg\varphi) = free(\varphi)$; $free(\varphi \wedge \psi) = free(\varphi) \cup free(\psi)$; $free(\exists x \varphi) = free(\varphi) - \{x\}$.

Con $\varphi(x_1, \ldots, x_n)$ denotamos una fórmula cuyas variables libres están incluidas en $\{x_1, \ldots, x_n\}$. Una *sentencia* es una fórmula sin variables libres. Si $\varphi$ es una fórmula atómica entonces $r_k(\varphi) = 0$, el $rk(\neg\varphi) = rk(\varphi)$, el $rk(\varphi \wedge \psi) = max\{rk(\varphi), rk(\psi)\}$ y el $rk(\exists x \varphi) = rk(\varphi) + 1$.

El fragmento la lógica de primer orden $FO^k$ se obtiene restringiendo $X$ a las variables $x_1, \ldots, x_k$.

El fragmento existencial $FO(\exists)$ consiste de aquellas fórmulas que se obtienen cerrando el conjunto de las formulas atómicas y las fórmulas atómicas negadas de $FO$ bajo las conectivas de conjunción y disjunción y de cuantificación existencial. $FO^k(\exists)$ es el fragmento con $k$ variables de $FO(\exists)$. La lógica $L_{\infty,\omega}$ es la extensión infinitaria usual de $FO$ la cual permite conjunciones y disjunciones sobre conjuntos arbitrarios de fórmulas. $L^k_{\infty,\omega}$ es el fragmento con $k$ variables de $L_{\infty,\omega}$. La lógica $L_{\infty,\omega}(\exists)$ es la extensión infinitaria de $FO(\exists)$ la cual permite conjunciones y disjunciones sobre conjuntos arbitrarios de fórmulas. $L^k_{\infty,\omega}(\exists)$ es el fragmento con $k$ variables de $L_{\infty,\omega}(\exists)$.

Un ejemplo para $\sigma = \langle E^2 \rangle$ en $FO^3(\exists)$ que obtiene los caminos de longitud $n$ es

$$\begin{aligned}
&\rho_1(x, y) = E(x, y) \\
&\rho_n(x, y) = \exists z(E(x, z) \wedge \exists x(x = z \wedge \rho_{n-1}(x, y)))
\end{aligned}$$

Un ejemplo para $\sigma = \langle E^2 \rangle$ en $L^3_{\infty,\omega}(\exists)$ que obtiene la clausura transitiva del grafo $E$ es

$$\varphi(x, y) = \bigvee_{n \geq 1} \rho_n(x, y)$$

## 2.2. Semántica de la Lógica de Primer Orden

Una *estructura* $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_s^{\mathcal{A}} \rangle$ de vocabulario $\sigma = \langle R_1, \ldots, R_s \rangle$ con aridades $r_1, \ldots, r_s$, respectivamente, consiste de un conjunto no vacío $A$, llamado universo o dominio de $\mathcal{A}$ y de una relación $R_i^{\mathcal{A}} \subseteq A^{r_i}$ para cada símbolo de relación $R_i$ de $\sigma$. En teoría de modelos finitos sólo consideramos estructuras con dominio finito.

En el ámbito de la teoría de bases de datos relacionales, un esquema de base de datos es equivalente a la noción de vocabulario relacional en el ámbito de teoría de modelos finitos. Si $\sigma = \langle R_1, \ldots, R_s \rangle$ es un esquema con aridades $r_1, \ldots, r_s$, respectivamente, una *base de datos* sobre el esquema $\sigma$ es una estructura $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_s^{\mathcal{A}} \rangle$. A veces escribiremos

$dom(\mathcal{A})$ en lugar de $A$. Nosotros denotaremos con $\mathsf{B}_\sigma$ a la clase de todas las bases de datos de esquema $\sigma$.

En adelante utilizaremos las palabras base de datos relacional y, estructura relacional finita, indistintamente. A una estructura de vocabulario $\sigma$ la denotamos como $\sigma$-estructura. Dadas dos bases de datos $\mathcal{A}$ y $\mathcal{B}$, con $\mathcal{A} \cong \mathcal{B}$ señalamos la existencia de un isomorfismo entre las bases de datos, es decir la existencia de una biyección, $f : dom(\mathcal{A}) \to dom(\mathcal{B})$ tal que para todo tupla $(a_1, \dots, a_{r_i}) \in A^{r_i}$, $(a_1, \dots, a_{r_i}) \in R_i^{\mathcal{A}}$ sii $(f(a_1), \dots, f(a_{r_i})) \in R_i^{\mathcal{B}}$.

En este contexto una *consulta* a una base de datos es una función que se aplica a una $\sigma$-estructura y produce como resultado una $\{R\}$-estructura con la aridad de $R$ igual a $r$, o un valor del conjunto $\{0,1\}$ en el caso que la consulta sea Booleana. Adicionalmente ésta debe preservar isomorfismos y ser computable. Notemos que el dominio de una consulta es $\mathsf{B}_\sigma$ y por lo tanto infinito.

**Definición 2.1 (Chandra y Harel 1980[2])** *Sean $\mathcal{A}$,$\mathcal{B}$ $\sigma$-estructuras y sea la consulta $r$-aria $q : \mathsf{B}_\sigma \to \mathsf{B}_{\{R^r\}}$ una función parcial. $q$ es una consulta computable sii*

1. *Para toda base de datos $\mathcal{A}$ de esquema $\sigma$ sobre la cual $q$ está definida, se cumple que $dom(q(\mathcal{A})) \subseteq dom(\mathcal{A})$.*
2. *$q$ es recursiva en alguna codificación lineal de las $\sigma$-estructuras.*
3. *para todo isomorfismo $h : dom(\mathcal{A}) \to dom(\mathcal{B})$ se cumple que o bien $q(\mathcal{B}) = h(q(\mathcal{A}))$, o $q$ está indefinida tanto en $\mathcal{A}$ como en $\mathcal{B}$.*

Antes de ver cómo una fórmula expresa una consulta, necesitamos definir la noción de satisfacción de una fórmula sobre una estructura o interpretación.

Sea $\mathcal{A}$ una $\sigma$-estructura. Una valuación en $\mathcal{A}$ es una función $V$ con dominio en el conjunto de las variables y valores en $dom(\mathcal{A})$, $V : \{x_1, x_2, \dots\} \to A$. Denotamos con $V(x/a)$ la valuación que coincide con $V$, excepto que $V(x/a)(x) = a$.

La relación de satisfacción, $\mathcal{A} \models \varphi[V]$ (la valuación $V$ satisface la fórmula $\varphi$ en $\mathcal{A}$), se define como sigue:

| | | |
|---|---|---|
| $\mathcal{A} \models t_1 = t_2[V]$ | sii | $V(t_1) = V(t_2)$ |
| $\mathcal{A} \models R(t_1, \dots, t_n)[V]$ | sii | $(V(t_1), \dots, V(t_n)) \in R^{\mathcal{A}}$ |
| $\mathcal{A} \models \neg\varphi[V]$ | sii | no es cierto $\mathcal{A} \models \varphi[V]$ |
| $\mathcal{A} \models \varphi \wedge \psi[V]$ | sii | $\mathcal{A} \models \varphi[V]$ y $\mathcal{A} \models \psi[V]$ |
| $\mathcal{A} \models \exists x\varphi[V]$ | sii | para algún $a \in A$, $\mathcal{A} \models \varphi[V(x/a)]$ |

Notemos que el valor de verdad de $\mathcal{A} \models \varphi[V]$ depende únicamente de los valores que la valuación $V$ asigna a las variables libres de $\varphi$. Es decir, si $V_1(x) = V_2(x)$ para toda variable $x$, libre en $\varphi$, luego $\mathcal{A} \models \varphi[V_1]$ sii $\mathcal{A} \models \varphi[V_2]$. Así si $\varphi = \varphi(x_1, \dots, x_n)$ y $V(x_1) = a_1, \dots, V(x_n) = a_n$ luego escribimos $\mathcal{A} \models \varphi[a_1, \dots, a_n]$ para indicar $\mathcal{A} \models \varphi[V]$. En este caso también diremos que la tupla $(a_1, \dots, a_n)$ satisface la propiedad $\varphi$. En particular si $\varphi$ es una sentencia, luego el valor de verdad de $\mathcal{A} \models \varphi[V]$ es independiente de $V$.

Supongamos que $\varphi(x_1, \dots, x_n)$ tiene exactamente a $\{x_1, \dots, x_n\}$ como variables libres y asumamos que están ordenadas de la siguiente forma, $x_1 \leq x_2 \leq \cdots \leq x_n$, entonces el conjunto de todas las valuaciones $V$, tales que $V(x_1) = a_1, \dots, V(x_n) = a_n$, que hacen verdadera a $\varphi(x_1, \dots, x_n)$ en la estructura $\mathcal{A}$ definen una relación sobre $A$.

$$\varphi^{\mathcal{A}} = \{(a_1, \dots, a_n) : a_1, \dots, a_n \in dom(\mathcal{A}) \wedge \mathcal{A} \models \varphi(x_1, \dots, x_n)[a_1, \dots, a_n]\}$$

Es decir $\varphi^{\mathcal{A}}$ es la relación definida por $\varphi$ sobre la estructura $\mathcal{A}$ y su aridad está determinada por el número de variables libres en $\varphi$. Formalmente diremos que una fórmula $\varphi(x_1, \ldots, x_n)$ de vocabulario $\sigma$, *expresa* una consulta $q$, si para toda base de datos $\mathcal{A}$ de esquema $\sigma$ $q(\mathcal{A}) = \varphi^{\mathcal{A}}$. Notemos que cuando usamos lógicas como lenguajes de consulta consideramos sólo las consultas totales. Ésto se debe a la semántica de Tarski que es la que usamos para definir la relación de satisfacción. De todos modos las *consultas parciales* pueden expresarse (como en [8]) utilizando un par de fórmulas donde la primera fórmula es una sentencia que define el dominio de la consulta.

## 2.3. Teoría de Modelos y Bases de Datos

Una referencia para su estudio es [1]. $\mathcal{L}_\sigma$ denotará la lógica $\mathcal{L}$ cuyas fórmulas bien formadas hacen mención a símbolos de relación del vocabulario $\sigma$. Dada una estructura $\mathcal{A}$ de vocabulario $\sigma$ la $\mathcal{L}$-*teoría* de $\mathcal{A}$ es el conjunto de sentencias de $\mathcal{L}$ que se satisfacen en $\mathcal{A}$, en símbolos:

$$Th_{\mathcal{L}}(\mathcal{A}) = \{\varphi \in \mathcal{L}_\sigma : \mathcal{A} \models \varphi\}$$

Es decir, la $Th_{\mathcal{L}}(\mathcal{A})$ son las consultas Booleanas que son verdaderas en la base de datos $\mathcal{A}$. La noción de equivalencia en $\mathcal{L}$, denotada por $\mathcal{A} \equiv_{\mathcal{L}} \mathcal{B}$, significa que las estructuras coinciden en sus $\mathcal{L}$-*teorías*. Es decir $\mathcal{A} \equiv_{\mathcal{L}} \mathcal{B}$ *sii* $Th_{\mathcal{L}}(\mathcal{A}) = Th_{\mathcal{L}}(\mathcal{B})$.

La noción de preservación de teorías en $\mathcal{L}$, denotado por $\mathcal{A} \preceq_{\mathcal{L}} \mathcal{B}$, significa que la $\mathcal{L}$-*teoría* de $\mathcal{A}$ está incluida en la $\mathcal{L}$-*teoría* de $\mathcal{B}$. Es decir $\mathcal{A} \preceq_{\mathcal{L}} \mathcal{B}$ *sii* $Th_{\mathcal{L}}(\mathcal{A}) \subseteq Th_{\mathcal{L}}(\mathcal{B})$.

Si $\varphi$ es una sentencia de $\mathcal{L}_\sigma$, definimos el conjunto de los modelos de $\varphi$ como:

$$MOD_{\mathcal{L}}(\varphi) = \{\mathcal{A} \in \mathsf{B}_\sigma : \mathcal{A} \models_{\mathcal{L}} \varphi\}$$

## 2.4. Tipos

Dada una base de datos $\mathcal{A}$ y una $k$-tupla $\bar{a} = (a_1, \ldots, a_k)$ sobre $\mathcal{A}$, nos interesa considerar todas las propiedades de $\bar{a}$ en la base de datos $\mathcal{A}$. Para ésto usamos la noción de *tipo*. Sea $\mathcal{L}$ una lógica y $\mathcal{A}$ una base de datos de algún esquema $\sigma$, el $\mathcal{L}$-*tipo* de $\bar{a}$ en $\mathcal{A}$, denotado por $tp^{\mathcal{L}}_{\mathcal{A}}(\bar{a})$, es el conjunto de fórmulas con variables libres entre $\{x_1, \ldots, x_k\}$ tal que cada fórmula del conjunto es verdadera cuando es interpretada en $\mathcal{A}$ bajo cualquier valuación que asigna el valor de la $i$-ésima componente $a_i$ de $\bar{a}$ a la variable $x_i$,para $1 \leq i \leq n$. En símbolos

$$tp^{\mathcal{L}}_{\mathcal{A}}(\bar{a}) = \{\varphi \in \mathcal{L}_\sigma : free(\varphi) \subseteq \{x_1, \ldots, x_k\} \wedge \mathcal{A} \models \varphi[a_1, \ldots, a_k]\}.$$

Destaquemos que, de acuerdo con esta definición, la $\mathcal{L}$-*teoría* de la base de datos $\mathcal{A}$, $Th_{\mathcal{L}}(\mathcal{A})$, está incluida en el $\mathcal{L}$-*tipo* de cada tupla sobre la base de datos $\mathcal{A}$.

También podemos considerar un $\mathcal{L}$-*tipo* sin tener en mente una base de datos particular. Es decir, adicionamos propiedades (fórmulas con la cantidad de variables libres adecuada) a un conjunto mientras éste permanezca consistente. Así podemos construir la clase de los tipos para todas las posibles $k$-tuplas sobre bases de datos de esquema $\sigma$. A esta clase la denotamos con $Tp^{\mathcal{L}}(\sigma, k)$. En símbolos

$$Tp^{\mathcal{L}}(\sigma, k) = \{tp^{\mathcal{L}}_{\mathcal{A}}(\bar{a}) : \mathcal{A} \in \mathsf{B}_\sigma \wedge \bar{a} = (a_1, \ldots, a_k) \in dom(\mathcal{A})^k\}$$

Sea $\alpha \in Tp^{\mathcal{L}}(\sigma, k)$ ($\alpha$ es el $\mathcal{L}$-*tipo* de alguna $k$-tupla sobre alguna base de datos en $\mathsf{B}_\sigma$), decimos que una base de datos $\mathcal{A}$ *realiza* el tipo $\alpha$ si existe una k-tupla $\bar{a}$ cuyo $\mathcal{L}$-*tipo* sea $\alpha$, es decir $tp^{\mathcal{L}}_{\mathcal{A}}(\bar{a}) = \alpha$. En símbolos

$$Tp^{\mathcal{L}}(\mathcal{A}, k) = \{tp^{\mathcal{L}}_{\mathcal{A}}(\bar{a}) : \bar{a} = (a_1, \ldots, a_k) \in dom(\mathcal{A})^k\}$$

El $\mathcal{L}$-*tipo* de una $k$-tupla $\bar{a}$ realizado en $\mathcal{A}$, representa la clase de todas las consultas expresables en $\mathcal{L}$ que cuando evaluadas en $\mathcal{A}$, contienen en su respuesta a $\bar{a}$ o a alguna de sus subtuplas, dependiendo de la aridad de la consulta. Informalmente si dos $k$-tuplas poseen el mismo $\mathcal{L}$-*tipo* entonces, o bien aparecerán las dos tuplas o ninguna de ellas en la respuesta de toda consulta expresable en $\mathcal{L}$.

## 2.5. Isomorfismos Parciales

Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, con $\sigma = \langle A, R_1, \ldots, R_n \rangle$, con $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_n^{\mathcal{A}} \rangle$ y con $\mathcal{B} = \langle B, R_1^{\mathcal{B}}, \ldots, R_n^{\mathcal{B}} \rangle$

$\mathcal{A}$ es ***subestructura inducida*** de $\mathcal{B}$ si y sólo si $dom(\mathcal{A}) \subseteq dom(\mathcal{B})$, $R_i^{\mathcal{A}} = R_i^{\mathcal{B}} \cap dom(\mathcal{A})^r$ con $1 \leq i \leq n$ y $r$ la aridad de $R_i$.

$\mathcal{A}$ es ***subestructura*** de $\mathcal{B}$ si y sólo si $dom(\mathcal{A}) \subseteq dom(\mathcal{B})$, $R_i^{\mathcal{A}} \subseteq R_i^{\mathcal{B}}$ con $1 \leq i \leq n$.

**Definición 2.2** *Sean $\mathcal{A}$ y $\mathcal{B}$ estructuras relacionales, sea $dom(p)$ y $rg(p)$ el dominio y rango de la función $p$ respectivamente, sea $p$ una función con $dom(p) \subseteq A$ y $rg(p) \subseteq B$, $p$ es un isomorfismo parcial si $p$ es un isomorfismo entre las subestructuras inducidas en $\mathcal{A}$ y $\mathcal{B}$ por $dom(p)$ y $rg(p)$ respectivamente.*

Los isomorfismos parciales preservan la validez de las fórmulas atómicas y sus combinaciones Booleanas. No así la validez de las fórmulas con cuantificadores. Para que ésto suceda es necesario que los isomorfismos parciales admitan extensiones.

# 3. Preservación de Teorías y Juegos de Fichas

Los juegos de fichas(pebble games) fueron introducidos por Ehrenfeucht y Fraïssé y son un medio algebraico para caracterizar la equivalencia elemental de estructuras, es decir la igualdad de teorías de $FO$. Se han realizado generalizaciones que caracterizan la preservación de teorías [5], es decir la inclusión de teorías, en lógicas menos expresivas que $FO$.

Los juegos también se pueden utilizar para caracterizar la relación que existe entre el *tipo* de una tupla $(a_1, \ldots, a_s)$ en la estructura $\mathcal{A}$ y el *tipo* de una tupla $(b_1, \ldots, b_s)$ en la estructura $\mathcal{B}$.

Notemos que la noción de preservación no siempre tiene sentido en cualquier lógica, por ejemplo la preservación de teorías en $FO$ implica la igualdad de teorías. Ésto es, supongamos que $Th(\mathcal{A}) \subseteq Th(\mathcal{B})$ y que existe una sentencia $\varphi$ tal que, $\varphi \in Th(\mathcal{B})$ y $\varphi \notin Th(\mathcal{A})$ entonces $\neg\varphi \in Th(\mathcal{A})$ y de aquí $\neg\varphi \in Th(\mathcal{B})$ lo que es absurdo. Ésto es así porque la teoría de una estructura siempre es consistente. Por lo tanto $Th(\mathcal{A}) = Th(\mathcal{B})$.

## 3.1. Preservación de Teorías en $FO^k(\exists)$

Los resultados expuestos en esta sección se deben a Kolaitis y Vardi, [5] y [6].

**Definición 3.1** *Sea $k \geq 1$, $n \geq 0$ y $\{(a_1, b_1), \ldots, (a_s, b_s), (a_{s+1}, a_{s+1}), \ldots, (a_k, a_k)\}$ un conjunto de pares de fichas. El juego $G_n^{FO^k(\exists)}(\mathcal{A}, a_1, \ldots, a_s, \mathcal{B}, b_1, \ldots, b_s)$ es un juego en el que participan dos jugadores, $JugadorI$ y $JugadorII$. El juego consiste de $n$ movimientos. Al inicio las fichas $a_1, \ldots, a_s$ han sido colocadas sobre $s$ elementos de $\mathcal{A}$ y las fichas $b_1, \ldots, b_s$ han sido colocadas sobre $s$ elementos de $\mathcal{B}$. En cada movimiento, el $JugadorI$ elige una de las fichas $a_i$, ya sea no usada si todavía quedan o una que se encuentra ya jugada, y la ubica sobre un elemento de la estructura $\mathcal{A}$. El $JugadorII$ responde ubicando la correspondiente ficha del par sobre un elemento de la estructura $\mathcal{B}$. El $jugadorII$ pierde si el conjunto $f = \{(a_1, b_1), \ldots, (a_i, b_i)\}$ no es un* **isomorfismo parcial** *entre $\mathcal{A}$ y $\mathcal{B}$. El $JugadorII$ gana el juego de $n$-movimientos si gana cada uno de los $n$ primeros movimientos.*

Informalmente el juego consiste en selecionar inicialmente una subestructura inducida por la tupla $\bar{a}$ sobre la estructura $\mathcal{A}$ y otra subestructura inducida por la tupla $\bar{b}$ sobre la estructura $\mathcal{B}$. Cada movimiento extiende la subestructura en $\mathcal{A}$ con un elemento del dominio. A esta extensión debe corresponderle una extensión de la subestructura en $\mathcal{B}$ que sea isomorfa a la primera. Cuando las subestructuras tienen tamaño $k$ se elimina uno de sus elementos y el correspondiente de la otra subestructura. Luego se realiza la extensión anterior, preservando el isomorfismo de las subestructuras. Notemos que el *$JugadorI$ siempre juega sobre la estructura $A$ y el $JugadorII$ sobre la $B$.*

Decimos que un *Jugador* tiene una estrategia ganadora en el juego $G_n^{FO^k(\exists)}(\mathcal{A}, a_1, \ldots, a_s, \mathcal{B}, b_1, \ldots, b_s)$, si gana independientemente de las elecciones del oponente.

**Definición 3.2** *Sea $\bar{a} = a_1, \ldots, a_s$. Escribimos $\mathcal{A}, \bar{a} \sim_n^{FO^k(\exists)} \mathcal{B}, \bar{b}$ sii el $JugadorII$ tiene una estrategia ganadora en el juego $G_n^{FO^k(\exists)}(\mathcal{A}, \bar{a}, \mathcal{B}, \bar{b})$*

En lo que sigue escribiremos $\mathcal{A}, \bar{a} \sim_n \mathcal{B}, \bar{b}$ en lugar de $\mathcal{A}, \bar{a} \sim_n^{FO^k(\exists)} \mathcal{B}, \bar{b}$ salvo que no quede claro desde el contexto.

Denotamos con $\bar{a}_i^a$ a la tupla que se obtiene reemplazando el elemento de la posición $i$, es decir la tupla $(a_1, \ldots, a_{i-1}, a, a_{i+1}, \ldots, a_s)$

El juego caracteriza si la tupla $\bar{a}$ y la tupla $\bar{b}$ satisfacen las mismas fórmulas de $FO^k(\exists)$ de rango cuantificacional menor o igual que $n$.

**Teorema 3.1** *Dadas $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $k \geq 1$, $\bar{a} \in A^s$ y $\bar{b} \in B^s$, $n, 0 \leq s \leq k$ y $\varphi(\bar{x}) \in FO^k(\exists)$ los siguientes son equivalentes:*

1. $\mathcal{A}, \bar{a} \sim_n^{FO^k(\exists)} \mathcal{B}, \bar{b}$
2. *Sea $rk(\varphi(\bar{x})) \leq n$, se cumple*
   *$\mathcal{A} \models \varphi(\bar{x})[\bar{a}]$ entonces $\mathcal{B} \models \varphi(\bar{x})[\bar{b}]$.*

**Corolario 3.1** *Dadas $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $k \geq 1$, $\bar{a} \in A^s$ y $\bar{b} \in B^s$, $0 \leq s \leq k$, $n \geq 0$, y $\varphi(\bar{x}) \in FO^k(\exists)$ los siguientes son equivalentes:*

1. $\mathcal{A}, \bar{a} \sim_n^{FO^k(\exists)} \mathcal{B}, \bar{b}$ *para todo* $n$
2. $\mathcal{A} \models \varphi(\bar{x})[\bar{a}]$ *entonces* $\mathcal{B} \models \varphi(\bar{x})[\bar{b}]$.

**Definición 3.3** *Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $\mathcal{L}$ una lógica, $\mathcal{A} \preceq^{\mathcal{L}} \mathcal{B}$ sii $Th^{\mathcal{L}}(\mathcal{A}) \subseteq Th^{\mathcal{L}}(\mathcal{B})$.*

**Corolario 3.2** *Dadas $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $k \geq 1$, los siguientes son equivalentes:*

1. $\mathcal{A} \sim_n^{FO^k(\exists)} \mathcal{B}$ *para todo* $n$
2. $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$.

Es decir la $FO^k(\exists)$ teoría de $\mathcal{A}$ se preserva en $\mathcal{B}$ si y sólo si hay una estrategia ganadora en el juego $G_n^{FO^k(\exists)}(\mathcal{A}, \mathcal{B})$ para todo $n$.

# 4. Preservación de Teorías y realización de tipos

Recordemos que el $\mathcal{L}$-tipo de una $k$-tupla es el conjunto de fórmulas con a lo sumo $k$ variables libres en $\mathcal{L}$ que satisface la $k$-tupla o cualquiera de sus subtuplas. En particular las propiedades de la subtupla vacía forman la $\mathcal{L}$ teoría de una estructura. Así la $\mathcal{L}$ teoría de una estructura $\mathcal{A}$ está incluida en el $\mathcal{L}$ tipo de toda $k$-tupla sobre $\mathcal{A}$. Es decir la clase de todas las propiedades de una $k$-tupla expresables en $\mathcal{L}$, incluye no sólo las propiedades de todas sus subtuplas sino también las propiedades de la estructura en si misma.

Recordemos que $Tp^{\mathcal{L}}(\mathcal{A}, k)$ es la clase de todos los $\mathcal{L}$ tipos para $k$-tuplas los cuales son realizados en $\mathcal{A}$. Es decir si $\alpha \in Tp^{\mathcal{L}}(\mathcal{A}, k)$ significa que existe una $k$ tupla $\bar{a}$ sobre $\mathcal{A}$ tal que $tp_{\mathcal{A}}^{\mathcal{L}}(\bar{a}) = \alpha$.

El siguiente resultado debido a Kolaitis y Vardi [6] establece que para estructuras finitas la igualdad de $L_{\infty\omega}^k$ *-teorías* es equivalente a la igualdad de $FO^k$*-teorías*. Es decir si dos bases de datos no son distinguibles por ninguna consulta booleana en $FO^k$, tampoco lo son aunque permitamos fórmulas infinitarias (ésto se debe a que se trata de bases de datos finitas).

**Proposición 4.1** *Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras finitas, $k \geq 1$, los siguientes son equivalentes:*

1. *Para cada sentencia* $\phi \in L_{\infty\omega}^k$, $\mathcal{A} \models \phi$ *sii* $\mathcal{B} \models \phi$
2. *Para cada sentencia* $\phi \in FO^k$, $\mathcal{A} \models \phi$ *sii* $\mathcal{B} \models \phi$.

Un resultado similar [7], se tiene para el fragmento infinitario existencial con $k$ variables, $L_{\infty\omega}^k(\exists)$.

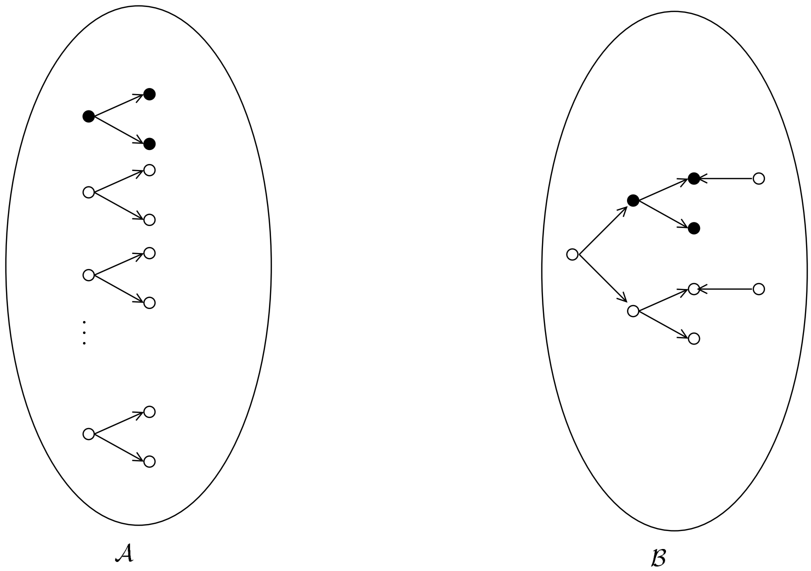
**Proposición 4.2** *Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras finitas, $k \geq 1$, los siguientes son equivalentes:*

1. *Para cada sentencia* $\phi \in L_{\infty\omega}^k(\exists)$, *Si* $\mathcal{A} \models \phi$ *entonces* $\mathcal{B} \models \phi$
2. *Para cada sentencia* $\phi \in FO^k(\exists)$, *Si* $\mathcal{A} \models \phi$ *entonces* $\mathcal{B} \models \phi$.

Es decir la teoría existencial infinitaria con $k$ variables de $\mathcal{A}$ se preserva en $\mathcal{B}$ si y sólo si la teoría existencial de primer orden con $k$ variables de $\mathcal{A}$ se preserva en $\mathcal{B}$.

La definición 4.1 establece una nueva relación entre los tipos realizados en dos bases de datos $\mathcal{A}$ y $\mathcal{B}$. Establece que para todo tipo $\alpha$ realizado en la estructura $\mathcal{A}$ debe existir un tipo $\beta$ de la estructura $\mathcal{B}$ que lo incluya. Es decir si consideramos que $\bar{a}$ es una tupla cuyo tipo es $\alpha$ entonces existe una tupla $\bar{b}$ de tipo $\beta$ tal que la subestructura inducida por $\bar{a}$ en $\mathcal{A}$ es isomorfa a la subestructura inducida por la tupla $\bar{b}$ en $\mathcal{B}$. Y para cada extensión $\bar{a}_{i}^{a}$ de $\bar{a}$ existe una extensión $\bar{b}_{i}^{b}$ de $\bar{b}$ tal que las respectivas subestructuras inducidas son isomorfas y así sucesivamente. En terminos de juegos de fichas significa que para cada movimiento en la estructura $\mathcal{A}$ podemos realizar uno en la estructura $\mathcal{B}$ tal que el $tp_{\mathcal{A}}^{FO^k(\exists)}(\bar{a}_{i}^{a}) \subseteq tp_{\mathcal{B}}^{FO^k(\exists)}(\bar{b}_{i}^{b})$.

Un ejemplo para $FO^3(\exists)$ y $\sigma = \langle E^2 \rangle$ es:




Note que la terna resaltada en la estructura $\mathcal{B}$ incluye estrictamente las propiedades de la terna resaltada en la estructura $\mathcal{A}$.

**Definición 4.1** *Sean $\mathcal{A}$ y $\mathcal{B}$ estructuras de esquema $\sigma$, $k \geq 1$, denotamos con $Tp^{FO^k(\exists)}(\mathcal{A}, k) \sqsubseteq Tp^{FO^k(\exists)}(\mathcal{B}, k)$ el hecho de que para todo $\alpha \in Tp^{FO^k(\exists)}(\mathcal{A}, k)$ existe $\beta \in Tp^{FO^k(\exists)}(\mathcal{B}, k)$ tal que $\alpha \subseteq \beta$.*

**Proposición 4.3** *Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $k \geq 1$, los siguientes son equivalentes:*

1. $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$
2. $Tp^{FO^k(\exists)}(\mathcal{A}, k) \sqsubseteq Tp^{FO^k(\exists)}(\mathcal{B}, k)$.

Demostración

1) $\Rightarrow$ 2)

Supongamos que $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$ y $Tp^{FO^k(\exists)}(\mathcal{A},k) \not\sqsubseteq Tp^{FO^k(\exists)}(\mathcal{B},k)$. Entonces existe una $k$-tupla $\bar{a}$ tal que $tp^{FO^k(\exists)}(\bar{a}) = \alpha$ y no existe $\beta \in Tp^{FO^k(\exists)}(\mathcal{B},k)$ tal que $\alpha \subseteq \beta$. Sea $\alpha = \{\psi_1(\bar{x}), \dots\}$ con $\psi \in FO^k(\exists)$, entonces la fórmula

$$\phi = \exists \bar{x} \left( \bigwedge_{\psi \in \alpha} \psi(\bar{x}) \right) \in Th_{L^k_{\infty\omega}(\exists)}(\mathcal{A}).$$

Por 1) y Prop. 4.2, $\phi \in Th_{L^k_{\infty\omega}(\exists)}(\mathcal{B})$, entonces existe una tupla $\bar{b} \in B^k$ tal que

$$\mathcal{B} \models \left( \bigwedge_{\psi \in \alpha} \psi(\bar{x}) \right) [\bar{b}],$$

entonces para cada $\psi \in \alpha$, $\mathcal{B} \models \psi(\bar{x})[\bar{b}]$ entonces $\beta = tp^{FO^k(\exists)}(\bar{b}) \supseteq \alpha$. Lo que es absurdo.

2) $\Rightarrow$ 1)

Sea $\alpha \in Tp^{FO^k(\exists)}(\mathcal{A},k)$ por def. de tipo $Th_{FO^k(\exists)}(\mathcal{A}) \subseteq \alpha$. Por hip. existe un tipo $\beta \in Tp^{FO^k(\exists)}(\mathcal{B},k)$ tal que $\alpha \subseteq \beta$ y supongamos que $\beta$ es el tipo de la tupla $\bar{b}$. Por lo tanto $Th_{FO^k(\exists)}(\mathcal{A}) \subseteq \beta$. Dado que la teoría es un conjunto de sentencias, $Th_{FO^k(\exists)}(\mathcal{A}) \subseteq \{\varphi \in FO^k(\exists) : \mathcal{B} \models \varphi[\bar{b}]$ y $\varphi$ es una sentencia$\} \subseteq \beta$. Entonces $Th_{FO^k(\exists)}(\mathcal{A}) \subseteq Th_{FO^k(\exists)}(\mathcal{B}) \subseteq \beta$. De modo que $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$.

**Corolario 4.1** *Sean $\mathcal{A}$ y $\mathcal{B}$ $\sigma$-estructuras, $k \geq 1$, los siguientes son equivalentes:*

1. $\mathcal{A} \preceq^{L^k_{\infty,\omega}(\exists)} \mathcal{B}$.
2. $Tp^{FO^k(\exists)}(\mathcal{A},k) \sqsubseteq Tp^{FO^k(\exists)}(\mathcal{B},k)$.

Demostración: Por prop. 4.2 y prop. 4.3.

## 4.1. Un algoritmo polinomial para la $FO^k(\exists)$-preservación

Sea $\sigma = \langle R_1, \dots, R_n \rangle$ con aridades $r_1, \dots, r_n$ y $\mathcal{A}$, $\mathcal{B}$ $\sigma$-estructuras.

**<u>proc</u>** $\bar{a} \cong \bar{b}$

| Este procedimiento determina si el mapping $\bar{a} \mapsto \bar{b}$ es un isomorfismo |
|---|

**<u>for</u>** $1 \leq i < j \leq k$ **<u>do</u>**
    **<u>if</u>** $\bar{a}_i = \bar{a}_j \ \wedge$
      $\bar{b}_i \neq \bar{b}_j$ **<u>then</u>** $return(false)$ **<u>fi</u>**      No es función
**<u>od</u>**
**<u>for</u>** $1 \leq i < j \leq k$ **<u>do</u>**
    **<u>if</u>** $\bar{b}_i = \bar{b}_j \ \wedge$
      $\bar{a}_i \neq \bar{a}_j$ **<u>then</u>** $return(false)$ **<u>fi</u>**      No es inyectiva
**<u>od</u>**

| La suryectividad es trivial ya que $\bar{a}$ y $\bar{b}$ tienen la misma longitud |
|---|

**for** $i = 1$ **to** $n$ **do**

$X := \{(\bar{t}, \bar{u}) :\mid \bar{t} \mid = \mid \bar{u} \mid = r_i \ , \ t_j = a_h \wedge u_j = b_h \ , \ 1 \leq j \leq r_i \ , \ 1 \leq h \leq k\}$

| El conjunto $X$ está formado por todos los pares de $r_i$-tuplas que se pueden formar con elementos de $\bar{a}$ y los correspondientes elementos de $\bar{b}$ |
|---|

**for** $(\bar{t}, \bar{u}) \in X$ **do**
**if** $\bar{t} \in R_i^{\mathcal{A}} \wedge \bar{u} \notin R_i^{\mathcal{B}}$ **then** $return(false)$ **fi**
**if** $\bar{t} \notin R_i^{\mathcal{A}} \wedge \bar{u} \in R_i^{\mathcal{B}}$ **then** $return(false)$ **fi**
**od**
**od**
$return(true)$
.

El procedimiento $\cong$ determina si las tuplas $\bar{a}$ y $\bar{b}$ satisfacen las mismas fórmulas atómicas de $FO^k(\exists)$. $\bar{a} \ncong \bar{b}$ denota que el procemiento $\cong$ retorna $false$.

**proc** $S$

| Este procedimiento calcula una relación binaria $S$ sobre $k$-tuplas tal que $(\bar{a}, \bar{b}) \in S$ si el $tp_{\mathcal{A}}^{FO^k(\exists)}(\bar{a}) \nsubseteq tp_{\mathcal{B}}^{FO^k(\exists)}(\bar{b})$ |
|---|

$S_{m+1} := \{(\bar{a}, \bar{b}) : \bar{a} \ncong \bar{b}\}$ — $\boxed{\bar{a} \text{ y } \bar{b} \text{ no satisfacen las mismas fórmulas atómicas}}$
$S_m := \emptyset$
**while** $S_m \neq S_{m+1}$ **do**
$S_m := S_{m+1}$
**for** $(\bar{a}, \bar{b}) \notin S_{m+1}$ **do**
**if** $\exists i$ — $\boxed{\text{Determina la existencia de movimientos en los que el jugador II pierde}}$
$\exists a \in A$
$\forall b \in B$
$\bar{a}_i^a \ncong \bar{b}_i^b$ **then** $S_{m+1} := S_{m+1} \bigcup \{(\bar{a}, \bar{b})\}$ **fi** **od**
**od**
$return(S_{m+1})$
.
**begin**

| Este programa verifica $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$ |
|---|

$T := (A^k \times B^k) - S$ — $\boxed{(\bar{a}, \bar{b}) \in T \text{ si el } tp_{\mathcal{A}}^{FO^k(\exists)}(\bar{a}) \subseteq tp_{\mathcal{B}}^{FO^k(\exists)}(\bar{b})}$
**if** $\forall \bar{a} \in A^k$
$\exists \bar{b} \in B^k$
$(\bar{a}, \bar{b}) \in T$ **then** $return(true)$ **else** $return(false)$
**fi**
**end**

El procedimiento $S$ determina qué pares de tuplas no satisfacen las mismas fórmulas de $FO^k(\exists)$. Inicialmente comienza con las tuplas que no satisfacen las mismas fórmulas atómicas de $FO^k(\exists)$. En en el paso $m$ de la iteración, considera las tuplas que hasta el momento satisfacen las mismas fórmulas de rango cuantificacional menor o igual que $m$, es decir las que todavía no están en $S$, y verifica si existe un movimiento de la ficha $\bar{a}_i$ tal que para todo movimiento de la ficha $\bar{b}_i$ el mapping $(\bar{a}^a_{\bar{i}} \mapsto \bar{b}^b_{\bar{i}})$ no es un isomorfismo. Es decir un movimiento en el que el $JugadorII$ pierde. En ese caso agrega el par de tuplas a $S$ ya que no satisfacen las mismas fórmulas de rango cuantificacional igual a $m+1$. Esta iteración se repite hasta que no se puedan agregar nuevos pares de tuplas a $S$. Note que el procedimiento $S$ siempre para ya que existen $\mid dom(\mathcal{A}) \mid^k \times \mid dom(\mathcal{B}) \mid^k$ pares distintos de tuplas. Finalmente nuestro algoritmo verifica la satisfacción de la definición 4.1 y por la proposición 4.3 se cumple que $\mathcal{A} \preceq^{FO^k(\exists)} \mathcal{B}$.

# 5. Otro Uso de la $FO^k(\exists)$-preservación

Con lo expuesto en las secciones anteriores también podemos concluir que la $FO^k(\exists)$-preservación indica que cada subestructura inducida por $k$ elementos del dominio de la base de datos $\mathcal{A}$ se encuentra inyectada en la base de datos $\mathcal{B}$. Esto sugiere que entre las bases de datos hay un cierto grado de similitud. Podemos relajar esta condición, y en lugar de exigir que toda $k$-tupla $\bar{a}$ (es decir la subestructura inducida por $\bar{a}$), se encuentre inyectada en $\mathcal{B}$ podemos pedir que un cierto número de ellas satisfaga la condición. Esto implicaría que ciertos objetos se comportan de manera similar en dos realidades distintas y así podemos concluir que nuestro algoritmo puede servir como estrategia para *data mining*.

# Referencias


[1] Abiteboul, S., Hull, R., Vianu, V., " Foundations of Databases", Addison-Wesley, 1994.

[2] Chandra, A. K., Harel D.: " Computable Queries for Relational Data Bases". Journal of Computer and System Sciences 21(2), pp. 156-178, 1980.

[3] Ebbinghaus, H. and Flum, J.: Finite Model Theory, 2nd ed. Springer (1999)

[4] Alejandro L. Grosso. " Preservación de Consultas a Bases de Datos Relacionales en Fragmentos Existenciales de la Lógica de Primer Orden". Tesis de Maestría. Univ. Nac. de San Luis. En preparación.

[5] Kolaitis, Phokion G. and Vardi, Moshe Y. On the Expressive Power of Datalog. PODS ACM 1990.

[6] Kolaitis, Phokion G. and Vardi, Moshe Y. Infinitary logics and and 0-1 laws . Information and Computation, 1992.

[7] Rosen Eric, Weinstein Scott. " Preservation Theorems in Finite Model Theory". Logic and Comp. Complexity, Lecture Notes in Computers Science 1994.

[8] Turull Torres, J.M. " On the expressibility and Computability of Untyped Queries", Annals of Pure and Applied Logic, vol 108, 1-3, p. 345-371, 2001.